A novel approach of longitudinal adverse event evaluation in oncology clinical trials: the Toxicity over Time (ToxT) analysis of clinical trials N9741 and 979254 (Alliance)

Gita Thanarajasingam, MD[1], **Pamela J. Atherton, MS**[2], **Paul J. Novotny, MS**[2], **Charles L. Loprinzi, MD (Professor of Oncology)**[1], **Jeff A. Sloan, Ph.D (Professor of Biostatistics and Oncology)**[2], and **Axel Grothey, MD (Professor of Oncology)**[1]

[1]Department of Medical Oncology, Mayo Clinic, Rochester, MN 55905

[2]Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN 55905

Abstract

Background—Traditional methods of reporting adverse events (AEs) in clinical trials are inadequate for modern oncology therapies with chronic administration. Conventional analysis and display of maximum grade AEs do not capture toxicity profiles that evolve over time or longer lasting, lower grade toxicity as does this longitudinal Toxicity over Time (ToxT) approach.

Methods—Graphical and analytical routines were compiled into an automated and standardized format to comprehensively analyze AEs. Plots visualizing summary statistics or individual patient data over discreet time points were combined with statistical methodology including longitudinal techniques (repeated measures models that describe the changes in AEs over each time period; time-to-event analyses of first, worst, or high grade; and area under the curve (AUC) analyses summarizing AE profiles over the entire study). The analytic capability of ToxT was demonstrated using two completed North Central Cancer Treatment Group (NCCTG)/Alliance clinical trials in cancer therapy (N9741, NCT00003594) and symptom control (979254).

Findings—Bar charts and stream plots showed higher incidences of dry mouth occurring late in 979254 for venlafaxine compared to placebo (week 1 [baseline]: 13% vs 22%, p=0.20; week 5: 49% vs 2%, p<0.0001) and increased nausea early for IROX vs FOLFOX in N9741 (cycle 1: mean grade 1.1 versus 0.6, p<0.0001). Event charts visually depicted earlier occurrences of higher

Corresponding author: Gita Thanarajasingam, MD, Mayo Clinic Rochester, 200 First Street SW, Rochester, MN 55905, Phone: +011-507-284-2511, Fax: +011-507-266-4972, thanarajasingam.gita@mayo.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

diarrhea grades for IROX patients and the AUC analysis indicated a higher magnitude of diarrhea experience over time in IROX compared to FOLFOX (4.2 versus 2.9, p<0.0001).

Interpretation—The ToxT analytic approach incorporates the dimension of time and offers a more comprehensive depiction of toxicity than current methods. With new, continuously administered targeted agents and maintenance regimens, these improved longitudinal analyses are directly relevant to patients and are imperative in oncology clinical trials.

Keywords

BACKGROUND

A consensus system for reporting of adverse events (AEs) is a cornerstone of clinical trials in oncology. Precise, complete and unbiased reporting of AEs is required to ensure the safety and tolerability of novel agents or combinations in cancer therapy trials. AE characterization is also important to patients and clinicians engaged in shared decision-making about a treatment strategy. There have been numerous initiatives to improve the quality of harms-related data reporting and to standardize toxicity reporting[1–4]. However, there has been little attention to modernizing methods of toxicity analysis so that they reflect current oncology therapeutics and trials.

Over the past decade, rapid expansion of novel, individualized therapies against cancer has driven a change in the complexity of the clinical trials investigating these drugs. Newer agents, such as targeted therapies and immunotherapy, are sometimes used continuously over months or years, rather than for a set number of cycles. Maintenance regimens are increasingly relevant in a variety of settings, from multiple myeloma post-transplant to metastatic colorectal cancer. Moreover, improvements in supportive care have facilitated extended durations of therapy. The consensus methodology for reporting AEs has not evolved in parallel with extended treatment durations.

There are significant limitations to current methods for capturing and displaying AE data. Tables of high-grade events as defined by the Common Terminology Criteria for Adverse Events (CTCAE)[5] traditionally display high grade events experienced during the entire trial. These analyses fail to define important information on when an AE will arise, its duration or its severity at a given point during therapy[6]. Importantly, conventional methods do not account for longer lasting, lower grade toxicities that may have substantial ramifications on quality of life. For example, an isolated episode of high grade diarrhea, whether or not causally associated with a study drug, is recorded, but chronic grade 2 diarrhea occurring daily over months at significant expense to a patient's quality of life gets lost in the toxicity assessment.

Inclusion of time-related information would provide a more comprehensive depiction of AEs that evolve over time. Alternate methods of longitudinal and graphical AE evaluation do

exist [7–11]. Some propose unique methods of summarizing AEs including bar charts and stream plots[12], but they do not focus on the comprehensive identification of patterns and differences in toxicity over time. Importantly, prior approaches have not been applied in an intuitive, clinically-oriented format and they have not been used for evaluation by regulatory agencies. We developed an analytic approach and standardized, comprehensive format, the Toxicity over Time (ToxT), which combines graphs and AE tabular displays with multiple longitudinal statistical techniques into a readily applicable tool for toxicity evaluation. In the current study, ToxT analysis is applied to data from two previously conducted cancer clinical trials at Mayo Clinic to demonstrate its utility for depicting AE profiles over time.

METHODS

Study Design

AE data from two completed North Central Cancer Treatment Group (NCCTG/Alliance) trials were used to demonstrate multiple longitudinal analyses that constitute the ToxT tool. N9741 was a randomized Phase III trial of combinations of oxaliplatin (OXAL), 5-fluorouracil (5-FU), and irinotecan (CPT-11) as initial treatment of metastatic colorectal cancer[13]. A total of 795 patients were randomly assigned to three treatment arms, of which we used the two most clinically relevant cohorts. FOLFOX (5-FU plus oxaliplatin) or IROX (CPT-11 plus oxaliplatin) chemotherapy were administered intravenously once every cycle. For FOLFOX one cycle was two weeks in length and for IROX one cycle was three weeks. Toxicity data was recorded at the end of each cycle. N9741 is registered with ClinicalTrials.gov, as NCT00003594. NCCTG/Alliance 979254 was a randomized Phase III trial comparing varying doses of venlafaxine to placebo in the management of hot flashes[14]. 191 patients had evaluable data for the study period. After a baseline assessment (week 1), participants took study medication orally for 4 weeks (weeks 2–5). AE data was collected at weekly intervals in this study. Statistical analyses could identify differences at each time point (per cycle on N9741 and per week on NCCTG 979254) and capture trends over time.

Procedures

Individual patient toxicity data on each trial was available for analysis through the NCCTG/Alliance database. Participants on both trials signed an IRB-approved, protocol-specific informed consent in accordance with federal and institutional guidelines.

Statistical Analysis

Data Analysis—Data analysis included testing of continuous data via Wilcoxon or t-tests and tests of discrete data via Chi-square or Fisher Exact tests to depict differences. Longitudinal techniques were also performed, and include repeated measures, time-to-event analysis using Kaplan-Meier (K-M) methodology, and area under the curve (AUC) analysis.

Repeated measures compares values collected at regular intervals between two treatment groups to assess if there are differences between arms over time. It takes into account within patient variations and between patient variations within the same treatment group, uses all available data per patient and does not exclude patients missing some AE data. Assumptions underlying the repeated measures model are tested for consideration of an alternative growth

model. Repeated measures considers the time variable to be discrete and thus time increments do not have to be the same length. In this model, the AE grade was entered as a continuous, dependent variable. The independent variables included age, sex, study arm, the cycle of treatment, and all of the stratification variables for the study. Age was entered as a continuous covariate, and all of the other variables were entered as nominal variables. The stratification variables are different for each study. The independent cycle variable was the repeated element in the model and it included the baseline cycle. Thus, the baseline AE is the starting point for the model. The input into ToxT currently assumes that all model assumptions (e.g. normality and homoscedasticity for analysis of variance testing) have been checked. A growth curve option exists for a continuous time variable, such as days or cycles of equal length. A significant p-value for the treatment group-time point interaction is an indication of differences in the profile over time.

K-M or Cox modeling are commonly applied to time-to-occurrence endpoints such as overall survival. These techniques are also useful when applied to time-to-AE occurrence since the incidence of an AE is important not only in its presence, but also in the time of onset.

Area Under the Curve (AUC) provides a single number to represent assessed values, such as the grades of diarrhea collected per cycle. The AUC is calculated with a mathematical formula that finds the area under a graphed line of data. The entire area or the pro-rated amount may be analyzed. Missing data are accounted for using various imputation techniques prior to the AUC analysis. AUC differences between two groups are compared using Kruskall-Wallis, Student's t-test or Wilcoxon test.

For all statistical procedures, distributional and procedural assumptions (for example, normality and homoscedasticity) are included in the algorithmic process. Data conversions for appropriate variance stabilizing transformations (square root, logarithmic) should be performed prior to applying the ToxT analysis. Further sensitivity analyses to account for missing data have been previously developed by our team[15, 16]. These include missing data algorithms involving 16 different types of assumptions for imputing missing data: single imputation, multiple imputation, minimum, maximum, mean and median imputation, last value carry forward, nearest neighbor analysis and Bayesian methods. These should also be utilized prior to applying ToxT analysis.

The ToxT tool was written using SAS software, version 9.4[17] and includes methodology for performing all of the analyses described above plus a macro (%Table)[18] developed by the Mayo Clinic Cancer Center Statistics statistical team to aid in analysis. The macro produces summary statistics and p-values for discrete and continuous data and contains an option for testing continuous data for normality prior to compiling the p-value, if needed. The macro can also produce K-M results.

Graphical Displays—AE data are displayed in many ways in the ToxT package. Bar charts display the frequency of events at a singular time point or over many time points. Stream plots display actual values or summary statistics (i.e. means) at specific time points connected by a line to show trends (i.e. slope) over time. Butterfly plots mirror data on both

sides of a central axis to visually compare individual patient data or summary statistics between two groups.

More dynamic graphs include K-M curves displaying time-to-event analyses and AUC plots, which are a visual representation of the area under the curve methodology described above. Event charts and heat maps also provide visualization of the AEs. Individual patient data is graphed resulting in an overall display of the AE profile of an entire cohort of patients over an entire trial. Event charts are created using a macro (%event) developed by the Mayo Clinic Cancer Center[10]. They provide unique patterns of toxicity data that can be grossly compared, and also have the capability to provide information to the resolution of an individual patient on study.

Role of the funding source

RESULTS

These results demonstrate the application of the ToxT analytic approach and its comprehensive, standardized outputs for longitudinal toxicity evaluation. The outputs below can be customized according to the needs of an individual study.

Output 1. Individual AEs at Each Time Point

A comparison of treatment groups may occur on a discrete interval (ex. weekly) or cycle-by-cycle basis to determine when AEs get worse, when they get better, and when they differ between arms. A weekly comparison of dry mouth incidence between venlafaxine 150 mg and placebo in patients on 979254 is shown in Table 1 (see appendix), which is produced using the %Table macro and could be constructed for each AE of interest on a clinical trial. It reveals a difference in this AE between the two cohorts beginning after week 1. The bar chart (Figure 1) serves as a graphical representation of N9741 data by each cycle and facilitates side-by-side comparison of diarrhea grade between the two regimens which could be readily reviewed with a patient. The difference in mean diarrhea grade between FOLFOX and IROX at cycle 1 was statistically significant (p<0.001) and FOLFOX is consistently seen to have less diarrhea than IROX.

Output 2. AE Trajectory Over Time

A stream plot shows AE trajectory in a format that is easily interpreted. An analysis using minutes or hours is applicable for events such as infusion reactions. Visualizing AEs over days might be appropriate for an AE such as delayed nausea and vomiting. Using weeks, cycles or months is appropriate for more chronic types of AEs such as chemotherapy-induced peripheral neuropathy.

The FOLFOX toxicity by cycle graph (Figure 2a) shows multiple treatment-related AEs over time in patients on N9741 in a single arm of the study. Nausea was experienced at a higher grade than the other AEs at cycle 1, then decreased over time. Other AEs such as stomatitis, remained steady over treatment. Clinically, these data suggest the role for augmented antiemetics early on in treatment with deescalation later on. A butterfly plot of these data uses the same y-axis for ready comparison of mean grades of each AE between two different treatment arms, displaying patients on FOLFOX (left) and IROX (right) (Figure 2b). The stream plot in Figure 2c is from NCCTG 979254, and reflects the mean dry mouth incidence in patients on venlafaxine versus placebo at weekly intervals. It is a visual display of the data shown in more detail on Table 1 (appendix). We see increased dry mouth on venlafaxine 150 mg compared to placebo, observed mainly in the later weeks of the study (incidence of 13% for venlafaxine and 22% for placebo at week 1 [baseline], p =0.20, versus 28% and 10%, respectively, at week 2, p=0.02, and a significant split with 49% with dry mouth on venlafaxine and 2% on placebo at week 5, p<0.0001). These data suggest that monitoring and management of this AE should begin around a month after initiation of drug. Defining the trajectory of an AE with this type of output is relevant to clinical trial design, as it may, for example, spur investigations of graduated dosing approaches if AEs are prevalent upfront.

Output 3. Repeated Measures for Rate of Change in AE Grades Over Time

A cycle is a continuous variable and therefore a growth curve model, a type of repeated measures analysis, is appropriate to compare two treatment arms over time. When displaying nausea grades between FOLFOX and IROX (Figure 3), the difference in arms over time is apparent. Nausea grades are higher for IROX than FOLFOX at each time point. At cycle 1 the mean grade was 1.1 for IROX versus 0.6 for FOLFOX (p<0.0001). This graph is a display of the mean nausea grade at each cycle for each arm and is not a result of the growth curve modeling. When applying growth curve methodology to these data, it was found that there is no difference in nausea trajectory over time (i.e. slope) between the two arms. The repeated measure result indicates a p-value of 0.18, which affirms no difference in the trajectory of nausea grades over time. Clinically, this might be due to a tachyphylaxis response where the symptom waned equally in both arms among most patients with repeated drug exposure. Alternatively, it might suggest that nausea on both arms was effectively managed with an antiemetic regimen. As a contrast, Figure 2c depicts a clear difference in slopes between arms (p<0.001).

Output 4: Time-to-Event Analyses for Onset of AEs

Time-to-event analysis provides information on the onset of AEs via graphic or tabular displays and associated p-values using K-M or Cox methodology. AE onset is clinically relevant in anticipating the need for an AE intervention. For example, the time to neutropenia suggests the potential need for growth factor. K-M analyses show that grade 2+ diarrhea onset occurs earlier for patients on IROX than those on FOLFOX in N9841 (Figure 4a). At 1 month, approximately 30% of patients receiving IROX experienced a grade 2 or higher diarrhea compared to less that 10% of those receiving FOLFOX, supporting an early role for increased antidiarrheals in irinotecan-containing regimens. The bar chart (Figure 4b) is alternate way to look at time-to-events. It depicts the time to first occurrence and time to

worst grade of six different AEs for the IROX patients. All but paresthesias had time to first occurrence within the first month of treatment, reflecting the gradual onset of neuropathy as compared to other AEs in oxaliplatin-based chemotherapy.

Output 5: Severity and Timing of AEs on an Individual Patient Level

Event charts and heat maps show overall patterns in AE data that can be visually compared between treatment arms. Each horizontal line in the event charts (figures 5a and 5b) for FOLFOX and IROX indicates a single patient's experience with diarrhea over time, thus offering resolution down to the level of an individual patient on a study. The graduated colors indicate the intensity of the AEs (white is no diarrhea, black is grade 4 or 5). The larger amount of dark color in the IROX portion indicates higher grades of diarrhea experienced on that arm early during treatment. Heatmaps produces similar graphics but shows gradual changing of the colors rather than grayscale. The high resolution of event charts and heatmaps provide a gross visual comparison of two cohorts and have the capability to identify subpopulations or individuals that may experience AEs differently. This could lead to investigation such as pharmacogenomics profiling of these patients.

Output 6: Area Under the Curve for Comparison of AEs over Time

An AUC analysis provides a single number to quantify numerous pieces of AE data over time and it captures low grade, longer lasting toxicity. In the conceptual example shown in Figure 6a below, Patient B has little toxicity except for an isolated grade 3 event, while patient A has consistent, grade 2 chronic toxicity over 7 cycles. Maximum grade analyses would depict Patient A's grade 3 experience and overlook Patient B's lower grade but chronic toxicity. However, the higher AUC accurately captures Patient B's substantial toxicity experience. Applied to the cohorts of patients on study N9741, the diarrhea grade over time had a mean AUC of 4.2 for IROX and 2.9 for FOLFOX ($p < 0.0001$) (Figure 6b). The AUC plot accounts for lower grade often subjective AEs such as chronic fatigue or dyspnea that have the potential to affect patient's quality of life. With a single value, the AUC depicts the entire course of treatment, compares an AE of interest between treatment arms readily and accounts for lower grade, longer lasting toxicity.

DISCUSSION

The evolution toward individualized medicine in oncology with new drugs and regimens that are often administered continuously has driven a shift in patients' experience of toxicity. Our methods for capturing and analyzing AEs in cancer trials should modernize to reflect these therapies. We developed the ToxT analysis, a novel approach to analyzing AEs that incorporates the dimension of time to provide a more complete, longitudinal depiction of chemotherapy toxicity than traditional methods.

ToxT analysis and its outputs can improve AE reporting in cancer trials. In applying the ToxT approach, we selected two very different trials, a cancer therapy trial (N9741) and a symptom control trial (979524), to demonstrate its versatility in defining the time profile of AEs. These trials were selected to introduce ToxT and its analyses, which are applicable to AEs from any chronically administered cancer therapy or supportive care. We are currently

applying this longitudinal AE analysis to trials of targeted agents in lymphoma and colorectal cancer at our institution. This approach can better quantify and qualify the impact of subjective AEs over time, such as the AUC of chronic grade 2 neuropathy, and it can also describe the time frame of objective AEs, for example time-to-neutropenia. While applying detailed ToxT analyses to every AE on a given study would not be feasible, we envision that AEs of interest or those with the highest frequency could be selected for longitudinal analysis. Assessing AEs over time may distinguish potentially overlapping toxicities and guide rational dosing approaches in trials. Furthermore, the ToxT offers high resolution analysis, to the level of the individual patient. This type of data could drive pharmacogenomic evaluation of patients with atypical AE responses. Longitudinal methods such as the AUC capture longer lasting, lower grade toxicity, which is of rising importance with chronically administered targeted agents, immunotherapy and maintenance strategies. Current methods overlook these toxicities which can substantially impact quality of life, and they should be accounted for in securing regulatory approval of novel agents with chronic administration.

In addition to its role in improving AE reporting in clinical trials, this analytic approach can directly impact patient care. Several of the graphic outputs can be reviewed directly with patients to facilitate a comprehensive understanding of the anticipated side effects of a given treatment. The between-regimen comparisons can be used to engage patients and oncologists in shared decision making about an upcoming treatment strategy. Additionally, identification of the time profile of an AE can allow for appropriate timing of supportive care when necessary, such as the use of antidiarrheals early on in treatment with irinotecan-based regimens. Demonstrating the time course of adverse events is valuable both to patients on study and off.

Widespread adoption of a novel, comprehensive toxicity analysis will require a paradigm shift. Maximum-grade toxicity reporting has been employed for decades and clinicians as well as study sponsors are accustomed to digesting its tabular format. Our method proposes significant changes to the way in which we process toxicity data, but ones that we feel are worthwhile given the benefits of longitudinal AE analysis in the current era of oncology therapeutics. In the past, systems for toxicity analysis had to be simple due to methodological limitations. The ToxT allows the use of current statistical technology to perform a broad range of analyses that delve deeper into data than ever before, particularly on the time profile of AEs.

Clinicians, principal investigators and regulatory agencies have obvious interest in longitudinal toxicity data, but statistical complexity, heterogeneity and lack of interpretability have limited broad application of existing longitudinal AE evaluation techniques[7, 12]. Accessibility and interpretability of the more complex AE analyses involved in the ToxT are limitations, but ones that can be addressed as we continue to develop and refine this tool. The sheer volume of toxicity-related data available to analyze in a longitudinal analysis is a potential barrier. For example, we recognize that information on AE management (such as timing of growth factors, antiemetics or antidiarrheals) is pertinent to the frequency of a given toxicity, and the lack of details on AE management are a limitation of this study. However, we elected to focus on the time profile of AEs and limit

the volume of data analyzed. It would be a valuable next step to collect data on the type and timing of supportive care measures to visualize how these affect toxicity over time, though this would contribute to the volume of data, which has a non-trivial effect on administrative and clerical burden for research assistants and principal investigators. Application of the methods presented and simplified in this report require significant biostatistician support to handle and refine the volume of toxicity data. However, the end outputs produced are intuitive to clinicians and clinical trialists, which is a strength of the ToxT approach. Furthermore, to reduce complexity, we do not anticipate that all of the AE analyses we have presented will be adopted by a given clinical practice, trial or regulatory body. Outputs would be selected and tailored based on the needs of a given study or patient population. We hope these analyses provide a foundation for a deeper understanding of toxicity and its mechanisms. As we move foward with development of this tool, we expect to modify it to include more succinct summarizations of AE data that may be performed without access to considerable biostatistician support.

Improving our statistical methods of AE analysis is the first step in modernizing our overall assessment of toxicity. Multiple other areas for improvement exist. Moving beyond the CTCAE criteria as they exist now and incorporating patent-reported outcomes is an opportunity to improve our current toxicity evaluation[19]. We can draw data directly from patients to accurately assess the experience of treatment with a given drug[20–22]. Furthermore, harnessing technology such as mobile devices and the internet to facilitate direct patient symptom reporting could provide the opportunity for real-time toxicity evaluation[23–25]. The novel approach of longitudinal AE analysis we have developed here can complement advances like these to ensure that the comprehensiveness and quality of toxicity data parallels the strides we continue to make in oncology therapy.

Acknowledgments

References

1. Peron J, Maillet D, Gan HK, Chen EX, You B. Adherence to CONSORT adverse event reporting guidelines in randomized clinical trials evaluating systemic cancer therapy: a systematic review. J Clin Oncol. 2013; 31(31):3957–63. [PubMed: 24062406]

2. Sivendran S, Latif A, McBride RB, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol. 2014; 32(2):83–9. [PubMed: 24323037]

3. Gwede CK, Johnson DJ, Daniels SS, Trotti A. Assessment of toxicity in cooperative oncology clinical trials: the long and short of it. J Oncol Manag. 2002; 11(2):15–21. [PubMed: 11989815]

4. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004; 141(10):781–8. [PubMed: 15545678]

5. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. Bethesda, MD: U.S. Department of Health and Human Services; 2009.

6. Thanarajasingam G, Hubbard JM, Sloan JA, Grothey A. The Imperative for a New Approach to Toxicity Analysis in Oncology Clinical Trials. Journal of the National Cancer Institute. 2015; 107(10)

7. Trotti A, Pajak TF, Gwede CK, et al. TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. Lancet Oncol. 2007; 8(7):613–24. [PubMed: 17543584]

8. Der, G.; Everitt, BS. Applied Medical Statistics Using SAS. Boca Raton, FL: CRC Press, Taylor & Francis Group, LLC; 2012.

9. Fitzmaurice, GM.; Laird, NM.; Ware, JH. Applied Longitudinal Analysis. 2. Hoboken, NJ: John Wiley & Sons, Inc; 2011.

10. Atherton PJ, Jasperson B, Nibbe A, et al. What Happened to All the Patients? Event Charts for Summarizing Individual Patient Data and Displaying Clinically Significant Changes in Quality of Life Data. Drug Information Journal. 2003; 37(1):11–21.

11. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. Pharmaceutical Statistics. 2008; 7(1):20–35. [PubMed: 17323410]

12. Yeh, S-T. Clinical Adverse Events Data Analysis and Visualization. The Pharmaceutical Industry SAS Users Group Conference Proceedings; 2007; Denver, Colorado: GlaxoSmithKline; 2007. p. Paper PO10

13. Goldberg RM, Sargent DJ, Morton RF, et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. J Clin Oncol. 2004; 22(1):23–30. [PubMed: 14665611]

14. Loprinzi CL, Kugler JW, Sloan JA, et al. Venlafaxine in management of hot flashes in survivors of breast cancer: a randomised controlled trial. Lancet. 2000; 356(9247):2059–63. [PubMed: 11145492]

15. Sloan JA, Dueck AC, Frost MH, et al. Applying QOL Assessments: Solutions for Oncology Clinical Practice and Research, Part 1. Current problems in cancer. 2005; 29(6):265–352.

16. Sloan JA, Dueck AC, Frost MH, et al. Applying QOL Assessments: Solutions for Oncology Clinical Practice and Research, Part 2. Current problems in cancer. 2006; 30(6):233–332.

17. SAS 9.4. Cary, North Carolina: SAS Institute Incorporated; 2013.

18. Novotny, PJ.; Tan, AD.; Foster, NR.; Sloan, JA. SAS Tools for Cost Effective and High Quality Clinical Trial Reporting. SAS Global Forum; 2012; Orlando, Florida. 2012.

19. Scharf O, Colevas AD. Adverse Event Reporting in Publications Compared With Sponsor Database for Cancer Clinical Trials. Journal of Clinical Oncology. 2006; 24(24):3933–8. [PubMed: 16921045]

20. Basch E, Abernethy AP, Mullins CD, et al. Recommendations for Incorporating Patient-Reported Outcomes Into Clinical Comparative Effectiveness Research in Adult Oncology. Journal of Clinical Oncology. 2012; 30(34):4249–55. [PubMed: 23071244]

21. Basch EM, Reeve BB, Mitchell SA, et al. Electronic toxicity monitoring and patient-reported outcomes. Cancer J. 2011; 17(4):231–4. [PubMed: 21799330]

22. Di Maio M, Gallo C, Leighl NB, et al. Symptomatic Toxicities Experienced During Anticancer Treatment: Agreement Between Patient and Physician Reporting in Three Randomized Trials. J Clin Oncol. 2015

23. Judson TJ, Bennett AV, Rogak LJ, et al. Feasibility of Long-Term Patient Self-Reporting of Toxicities From Home via the Internet During Routine Chemotherapy. Journal of Clinical Oncology. 2013; 31(20):2580–5. [PubMed: 23733753]

24. Basch E, Iasonos A, Barz A, et al. Long-Term Toxicity Monitoring via Electronic Patient-Reported Outcomes in Patients Receiving Chemotherapy. Journal of Clinical Oncology. 2007; 25(34):5374–80. [PubMed: 18048818]

25. Pietanza MC, Basch EM, Lash A, et al. Harnessing technology to improve clinical trials: study of real-time informatics to collect data, toxicities, image response assessments, and patient-reported outcomes in a phase II clinical trial. J Clin Oncol. 2013; 31(16):2004–9. [PubMed: 23630218]

RESEARCH INTO CONTEXT

Evidence before this study

Conventional reporting of toxicity in oncology trials is inadequate in the era of chronically administered, novel cancer therapies. The traditional maximum grade approach to toxicity reporting does not depict onset, duration or trajectory of adverse events, nor does it address longer lasting, lower grade toxicities that may occur at substantial expense to a patient's quality of life. Oral targeted agents, for example, are often administered daily, over several months or even years. Maintenance regimens are now relevant in a variety of settings, from myeloma post-transplant to metastatic colorectal cancer. Narrow focus on high grade toxicity is insufficient and potentially misleading. It does not reflect how severe nausea will be at cycle 2, whether a characteristic desquamative rash will occur days or weeks into therapy, or how many patients endure daily grade 2 diarrhea over months. This type of information is important to individual patients, imperative in clinical trials and may also bear significance in the process of securing regulatory approval for novel agents in the future. Longitudinal and graphical methods exist, but to our knowledge, there have not been any clinically focused efforts specifically aimed at modernizing the approach to adverse event evaluation to better reflect side effects of newer, chronic therapies in oncology. The current study aims to challenge conventional paradigms of AE reporting and present a novel approach to toxicity analysis that portrays adverse events over time.

Added value of this study

An improved, clinically oriented, longitudinal approach AE analysis fulfills an important and thus far unmet need in oncology. The current study is a groundbreaking endeavor to transform toxicity assessment in oncology clinical trials. We developed the Toxicity over Time (ToxT) analysis, an automated, standardized, longitudinal approach that constructs clinically meaningful statistical summaries of AE data over time. We demonstrate ToxT analyses with AE data from a completed cancer therapy trial and a symptom control trial. This approach has an role in the clinic, for optimally counseling individual patients on anticipated side effects of a given therapy. Longitudinal toxicity analysis is also critical in oncology trials, to better depict AEs of novel agents or combinations, and facilitate patient-centered clinical trials.

Implications of all the available evidence

This study demonstrates the practical application of a new, longitudinal approach to AE analysis, the Toxicity over Time. The variety of outputs demonstrated in the ToxT analysis uncover clinically relevant information such as time-to-onset of adverse events. Time-dependent toxicity data can spur further investigations - of graduated dosing approaches or the most appropriate timing of symptom control measures, for example. Some analyses, such as the event chart, offer the opportunity to identify subpopulations of patients with atypical AE responses for pharmacogenomic profiling or other individualized assessments. We feel that toxicity evaluation that includes the time profile of AEs in addition to their grade is more comprehensive and meaningful than conventional focus on grade alone. Longitudinal toxicity analysis is applicable and

important in a broad range of oncology studies and tumor types and has the potential to improve oncology clinical trials.
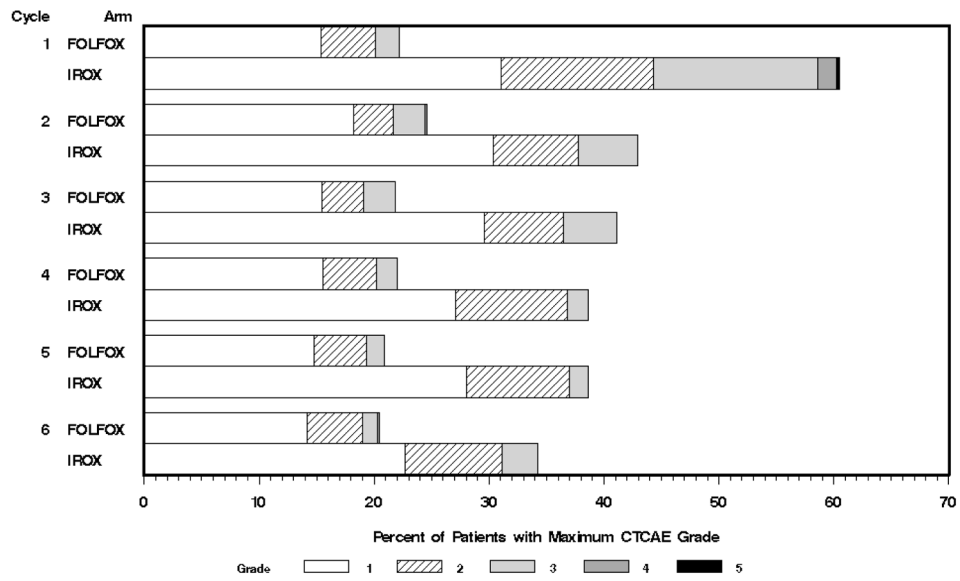
Figure 1. Diarrhea Incidence of FOLFOX versus IROX By Cycle and Grade
Output 1 demonstrates individual AEs at a given time point between different study arms.
The bar chart of diarrhea on FOLFOX and IROX by cycle and grade demonstrates the most substantial difference in diarrhea incidence between the two arms at cycle 1, with consistently less diarrhea on FOLFOX than IROX over time.

Figure 2a
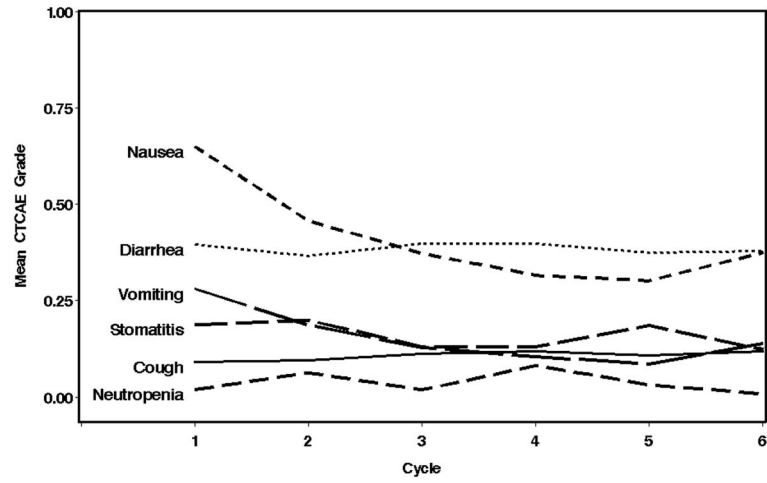
Figure 2b
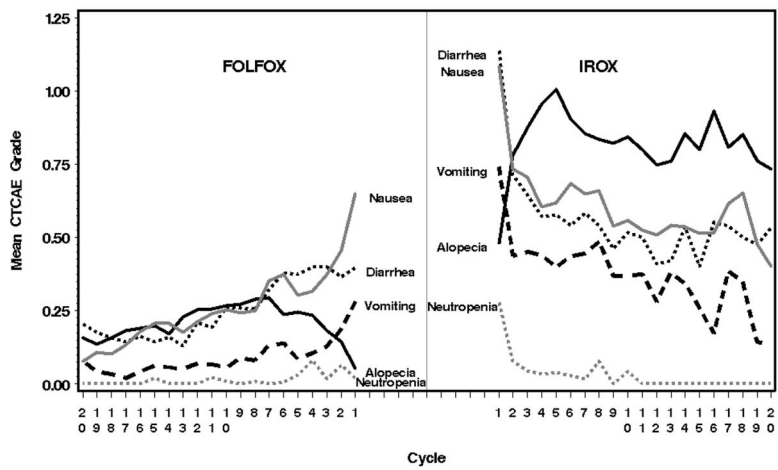
Figure 2a

Figure 2c
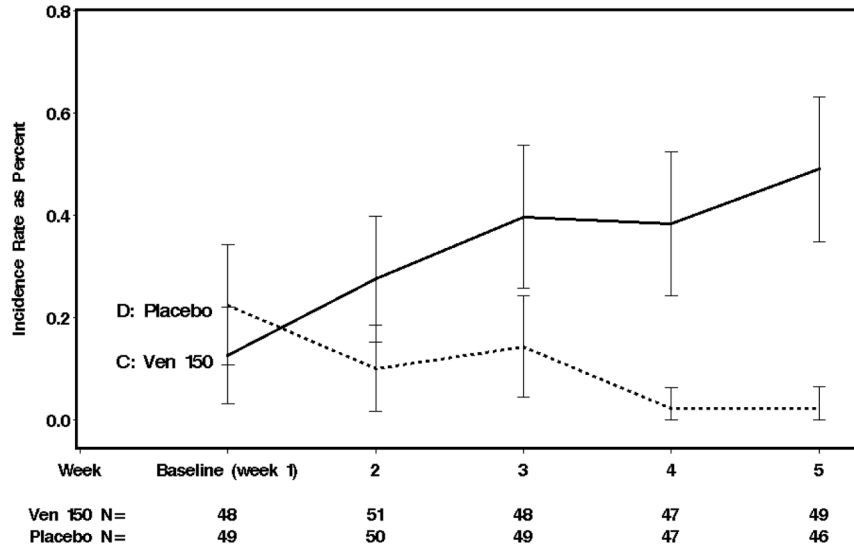
Figure 2b

Figure 2. AE Trajectory Over Time–Stream Plots

Output 2 highlights AE trajectory over time. Stream plot 2a shows several treatment-related AEs on a single study arm over time (cycles, in this study), and illustsrates the higher grade of gastrointestinal AEs compared to others early in FOLFOX treatment. The butterfly plot (2b) employs a shared y-axis to facilitate comparison of multiple AEs between two different treatment arms (FOLFOX versus IROX) over time. Stream plot 2c demonstrates comparison of a single AE of interest between two study arms to readily display the delayed time profile of dry mouth on venlafaxine versus placebo.

Figure 2a. Multiple Toxicities of FOLFOX By Cycle - Stream Plot

Figure 2b. Multiple Toxicities of FOLFOX and IROX By Cycle - Butterfly Plot

Figure 2c. Dry Mouth on Venlafaxine versus Placebo By Cycle

Arm by Cycle Interaction P-value= 0.18

Cycle	1	2	3	4	5	6
FOLFOX N=	660	587	520	449	367	286
IROX N=	368	337	286	248	207	172
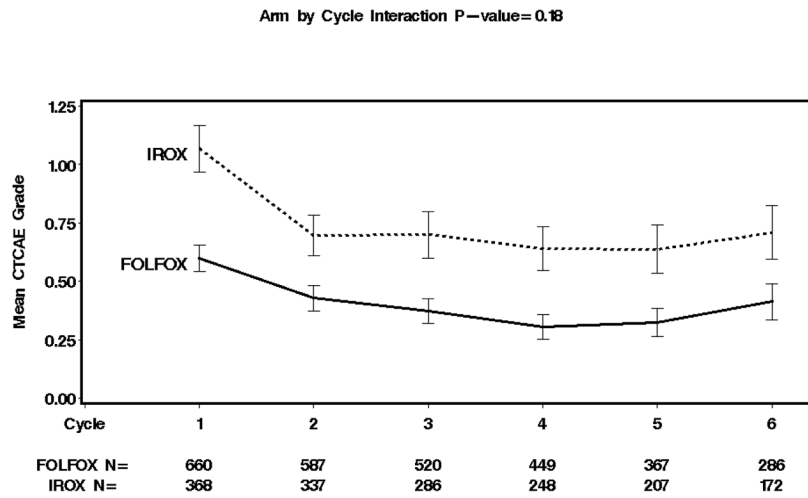
Figure 3. Nausea on FOLFOX versus IROX By Cycle with Repeated Measures
Output 3, repeated measures analysis, assesses the trajectory (slope) of an AE over time between arms. The p-value of 0.18 affirms no difference in th trajectory of nausea grades over time between FOLFOX and IROX.
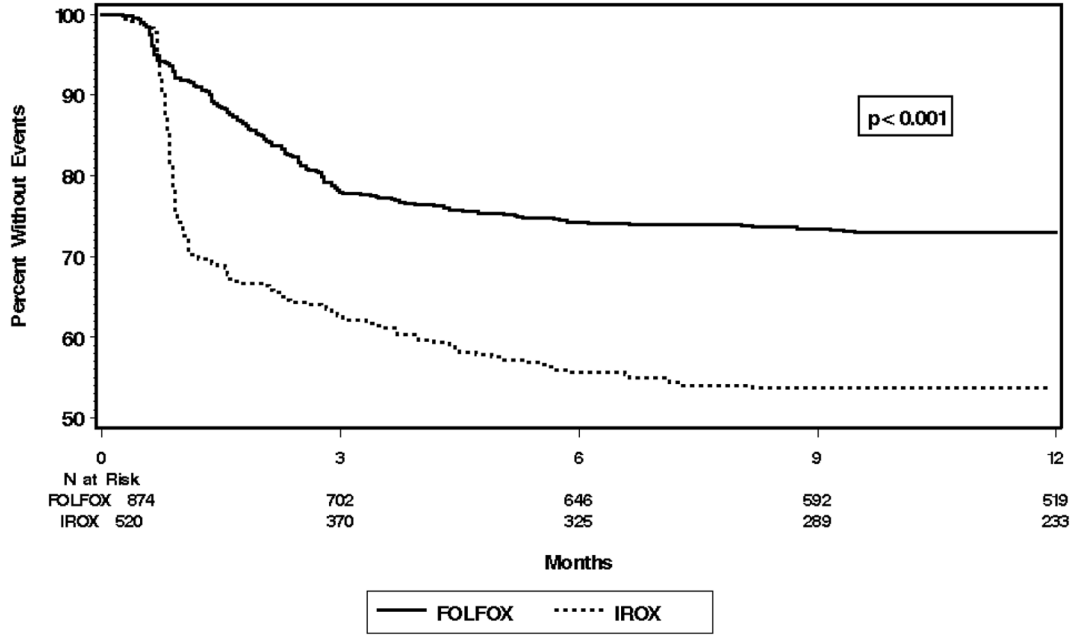
Figure 4a

Figure 4a

Figure 4b
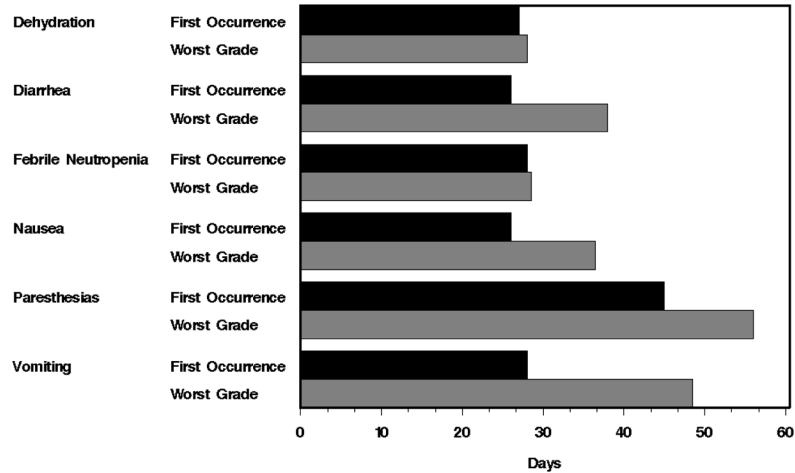
Figure 4b

Figure 4. Time-to-Event Analyses for Onset of AEs

Time-to-event analyses from Output 4 display information about the onset of AEs. The time to occurrence of a single AE between arms is highlighted in the Kaplan-Meier curve (Figure 4a), which depicts earlier time to Grade 2+ diarrhea on IROX. Figure 4b demonstrates time to first occurrence and worst grade of multiple AEs in a single arm, IROX. In this plot, the

more gradual onset of neuropathy compared to other AEs in oxaliplatin-based chemotherapy is seen.

Figure 4a. Time to Grade 2+ Diarrhea on FOLFOX versus IROX

Figure 4b. Median Time to First Occurrence and Worst Grade Toxicity on IROX
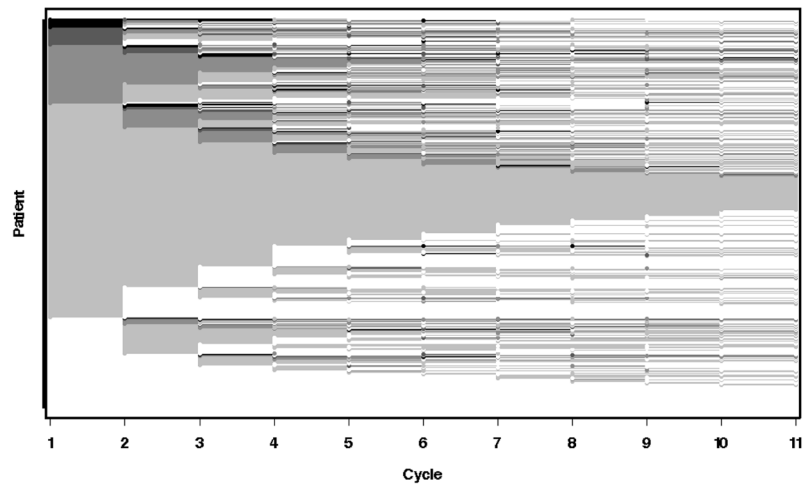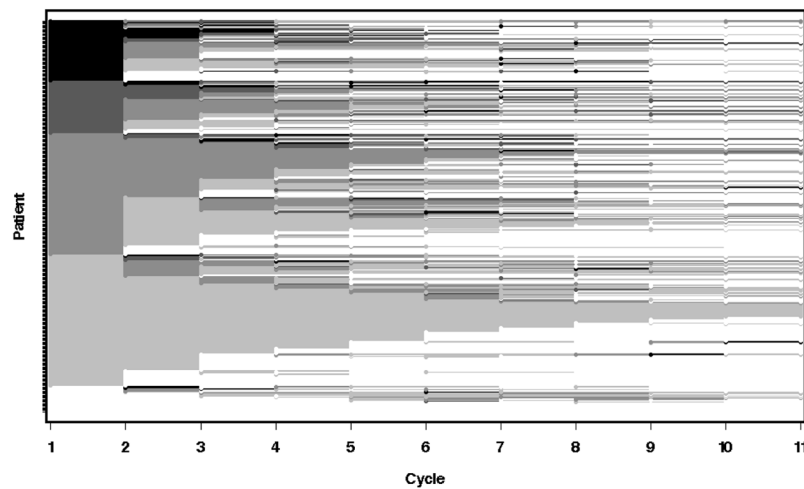
Figure 5a

Figure 5b

Figure 5. Event Charts for Severity and Timing of AEs on an Individual Patient Level
The event charts produced in Output 5 display overall patterns of AE data for gross comparison between two cohorts. Graduated colors indicate the AE grade (white is no diarrhea, black is grade 4 or 5) and readily display higher grade diarrhea earlier on in IROX (5b) than FOLFOX (5a). As each horitontal line indicates a single patient's experience with an AE over interest over time, event charts offer resolution down to the level of an individual study participant.
Figure 5a. Diarrhea on FOLFOX–Event Chart
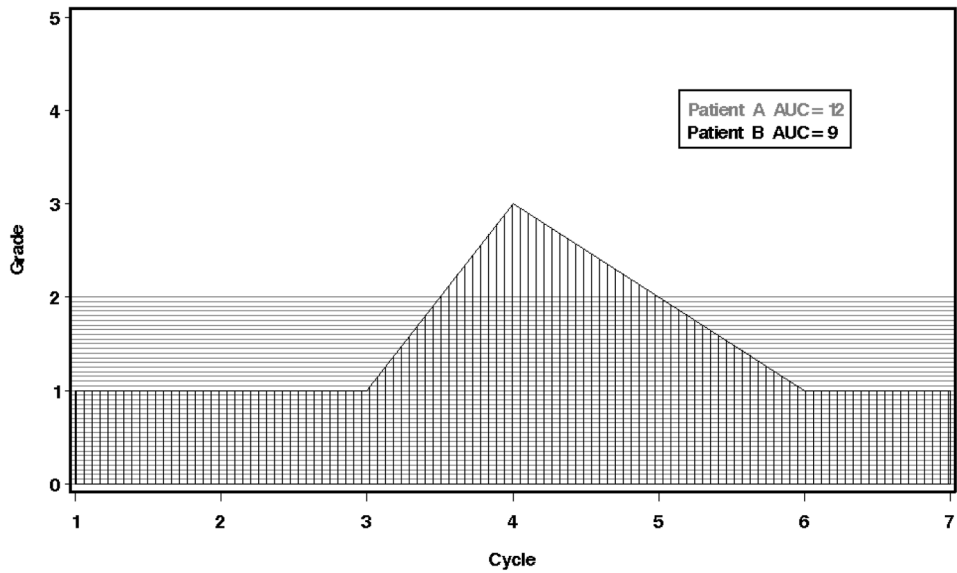Figure 5b. Diarrhea on IROX–Event Chart
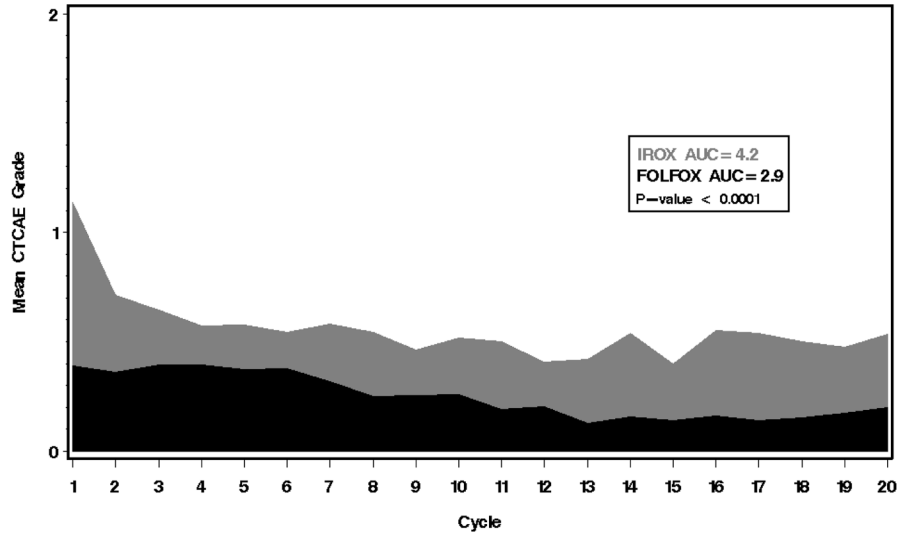
Figure 6a

Figure 6a

Figure 6b

Figure 6b

Figure 6. Area Under the Curve (AUC) Analysis to Compare AEs over Time

AUC analysis from Output 6 quantifies numerous pieces of AE data over time and accounts for low grade, longer lasting toxicity in a single number. In Figure 6a, patient A lacks any high grade event, but demonstrates chronic grade 2 toxicity that produces a higher AUC than Patient B, who has a single high grade event. The AUC captures the substantial toxicity

experience of low grade, longer lasting AEs. AUC analysis can be applied to entire cohorts on study (Figure 6b), in this case readily revealing a consistently higher grade diarrhea experience over time on IROX compared to FOLFOX.

Figure 6a. Conceptual Example of Area-Under-Curve Analysis

Figure 6b. Area Under Curve Analysis Applied–Diarrhea Grade Over Time on FOLFOX versus IROX

Table 1

Dry Mouth Frequency Per Week on Venlafaxine versus Placebo

In addition to generating bar charts, Output 1 produces tables that compare of AE incidence between arms at various time points. Table 1 below depicts dry mouth on venlafaxine compared to placebo by week on study, demonstrating a significant difference in arms that is most apparent in the later weeks of therapy.

	C: Ven 150 (N=54)	D: Placebo (N=56)	Total (N=110)	p value
Week 1 (Baseline):				0.20
Dry Mouth				
Missing	6	7	13	
No	42 (87.5%)	38 (77.6%)	80 (82.5%)	
Yes	6 (12.5%)	11 (22.4%)	17 (17.5%)	
Week 2: Dry Mouth				0.02
Missing	3	6	9	
No	37 (72.5%)	45 (90.0%)	82 (81.2%)	
Yes	14 (27.5%)	5 (10.0%)	19 (18.8%)	
Week 3: Dry Mouth				0.0049
Missing	6	7	13	
No	29 (60.4%)	42 (85.7%)	71 (73.2%)	
Yes	19 (39.6%)	7 (14.3%)	26 (26.8%)	
Week 4: Dry Mouth				<0.0001
Missing	7	9	16	
No	29 (61.7%)	46 (97.9%)	75 (79.8%)	
Yes	18 (38.3%)	1 (2.1%)	19 (20.2%)	
Week 5: Dry Mouth				<0.0001
Missing	5	10	15	
No	25 (51.0%)	45 (97.8%)	70 (73.7%)	
Yes	24 (49.0%)	1 (2.2%)	25 (26.3%)	